ISSN 2710-1673. Штучний інтелект, 2020, № 3

УДК 004.83 DOI: https://doi.org/10.15407/jai2020.03.045

Є.О. Купцова, С.К. Рамазанов
Київський національний економічний університет імені Вадима Гетьмана, Україна
пр. Перемоги, 54/1, м. Київ, 03057

АНАЛІЗ МОДЕЛЕЙ НАВЧАННЯ ШТУЧНИХ НЕЙРОННИХ МЕРЕЖ ДЛЯ ПЕРЕДБАЧЕННЯ ЦІН АВІАКВИТКІВ

E.A. Kuptsova, S.K. Ramazanov
Kyiv National Economic University named after Vadym Hetman, Ukraine
54/1, Peremogy Ave., Kyiv, 03057

ANALYSIS OF ARTIFICIAL NEURAL NETWORKS TRAINING MODELS FOR AIRFARE PRICE PREDICTION

Анотація. Авіаційний транспорт з кожним роком відіграє все більшу роль у світовій економіці. Цьому сприяють технологічний розвиток і новітні розробки в авіаційній галузі, глобалізація. У даній статті наведено огляд методів навчання штучних нейронних мереж для передбачення вартості авіаквитка. Було проаналізовано статті за 2017-2019 роки з метою визначення моделі з найточнішим результатом передбачень. Науковці проводили дослідження на відкритих даних, зібраних власноруч, та ставили собі за мету створити модель, яка б радила користувачу найкращий час для купівлі авіаквитка, коли ціна на нього була б найнижчою. У результаті огляду праць було виявлено, що найвищі результати мають моделі Bagging Regression Tree із точністю в 88% та метод випадкового лісу із точністю в 87%. Цивільна авіація відіграє важливу роль в економіці кожної держави. Авіація – найкращий засіб для подолання великих відстаней із комфортом за найкоротший час. Авіакомпанії пропонують клієнтам різноманітні можливості для подорожі як всередині країни, так і за її межами. Головною проблемою взаємодії авіакомпаній та клієнтів є вартість авіаквитків: перші хочуть продати побільше й подорожче, а другі – купити подешевше. Тому компанії використовують власні закриті алгоритми динамічного ціноутворення та постійно відслідковують ринкову ситуацію, швидко реагуючи на зміни попиту та дії конкурентів. Така поведінка дозволяє досягти балансу між бажаннями авіакомпаній та клієнтів. Науковці намагаються відшукати спосіб передбачення ціни на авіаквитки, аби авіамандрівники могли придбати їх за найнижчою ціною. Результати проведеної роботи у цій сфері складають загальні правила влучної покупки. Наприклад, згідно зі статтею (Udachny, 2016), для купівлі квитка через сайт expedia.com на внутрішній рейс у США найкращим днем є неділя, а найкращим періодом – за 57 днів до вильоту. У даній статті наведено огляд праць, автори яких порівнювали моделі машинного навчання. Досягнення даного напрямку роботи обмежуються лише прямими рейсами певного внутрішнього ринку (США, Індія) та 88% точності прогнозу (Tziridis et al., 2017). Найкращим результатом можна вважати Bagging Regression Tree-модель, описану в статті (Tziridis et al., 2017). Описана навчена модель може робити передбачення лише на основі двох параметрів: кількості безкоштовного вантажу та кількості днів, що залишились до вильоту, та має точність у 88%.

Ключові слова: машинне навчання; порівняльний аналіз; авіакомпанії; прогнозування цін.

Abstract. Air transport is playing an increasing role in the world economy every year. This is facilitated by technological development and the latest developments in the aviation industry, globalization. This paper provides an overview of artificial neural network training methods for airfare predicting. The articles for 2017-2019 were analyzed in order to determine the model with the most accurate prediction. The researchers conducted research on open data collected by themselves and set themselves the goal of creating a model that would advise a user the best time to buy a ticket when the price would be the lowest. The review of the papers by similar themes revealed that the Bagging Regression Tree model has the highest results with an accuracy of 88% and the random forest method has an accuracy of 87%. Civil aviation plays an important role in the economy of each country. Aviation is the best way to cover long distances in comfort in the shortest time. Airlines offer customers a variety of opportunities to travel both within the country and abroad. The main problem of interaction between airlines and customers is the airfare: the former want to sell more at the higher price, and the latter want to buy cheaper. Therefore, companies use their own private algorithms for dynamic pricing and constantly monitor the market situation, responsive to changes in demand and the actions of competitors. This behavior allows them to achieve a balance between the desires of airlines and customers. Scientists are trying to invent a way to predict airfare so that air travelers can buy them at the lowest price. The results of the work in this area provide general rules for the best purchase. For example, according to the article (Udachny, 2016) the

best day to buy a ticket by expedia.com for a domestic flight on the United States is Sunday, and the best period is 57 days before departure. This article provides an overview of the works, the authors of which compared the models of machine learning. Achievements in this area are limited to direct flights of a certain domestic market (USA, India) and 88% accuracy of the forecast (Tziridis et al., 2017). The Bagging Regression Tree model described in the article (Tziridis et al., 2017) can be considered the best result. This trained model can make predictions based only on two parameters: the number of free cargo and the number of days left before departure and has an accuracy of 88%.
Keywords: machine learning; comparative analysis; airlines; airfare prediction.

Вступ

Ціна авіаквитка складається з багатьох факторів: довжина перельоту, авіаперевізник, ціна на авіапаливо, день купівлі тощо. Кожна авіакомпанія, намагаючись отримати максимальний прибуток, використовує свій алгоритм динамічного ціноутворення. Для демонстрації взаємодії авіакомпаній та клієнтів у статті (Abdella et al., 2019) створено діаграму (рис. 1), що висвітлює ключові моменти: авіакомпанії хочуть продавати за найвищу ціну стільки квитків, скільки можливо, а клієнти хочуть купляти квитки за найнижчою ціною. Вирішуючи дану проблему, науковці намагаються за допомогою різних інструментів передбачення розробити найточніший алгоритм, однак авіакомпанії періодично оновлюють свої алгоритми, що ускладнює задачу прогнозування.

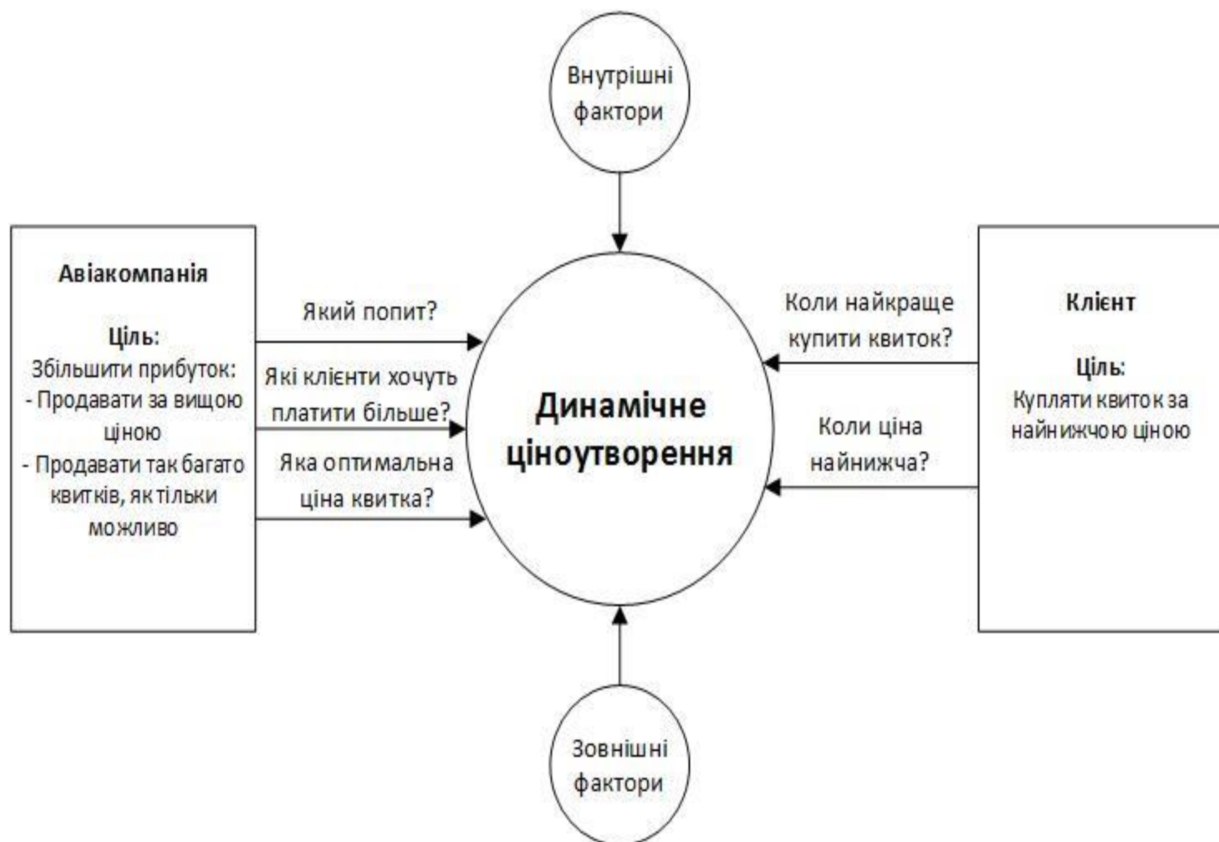


Рис. 1. Діаграма взаємодії авіакомпаній та клієнтів,
перекладена автором, розроблена авторами статті (Abdella et al., 2019)

Авіакомпанії намагаються забезпечити повне завантаження літака. Однак, зробити це – нелегка задача. Згідно з даними, наведеними у (Shy, 2008, с. 298), на ринку внутрішніх авіаперевезень США скасовуються 30 - 60% бронювань, а відсоток неявки на рейс сягає 8%. Із цією метою, зазвичай, продається квитків більше,

ніж є місць у літаку. Таке явище називається овербукінг (англ. *overbooking*) і воно покликане збільшити відсоток завантаженості літака.

Клієнти авіакомпаній, у свою чергу, намагаються знайти найдешевший авіаквиток на даний рейс. Існує думка, що найкраще купувати якомога раніше, хоча багато досліджень спростовують це. Проведений у статті (Groves & Gini, 2013) аналіз вказує, що стратегія «найраніше – найдешевше» лише зрідка виправдовує себе, оскільки часто авіакомпанії знижують ціни на день за тиждень до вильоту, а потім знову збільшують їх. На рис. 2 зображено графік цін будь-якої авіакомпанії на 5-денні рейси Нью Йорк – Міннеаполіс туди-назад із вильотом у четвер. Суцільною лінією вказані ціни на виліт (загалом, 8 рейсів).

Результати дослідження зі статті (Сериков, 2019) вказують, що, наприклад, на рейс Ростов-на-Дону – Санкт-Петербург (авіакомпанії «Азимут») за 20 годин до вильоту зафіксовано найнижчу ціну в 4090 руб., у той час як за 72 години до вильоту білет коштував 9390 руб., майже удвічі дорожче!

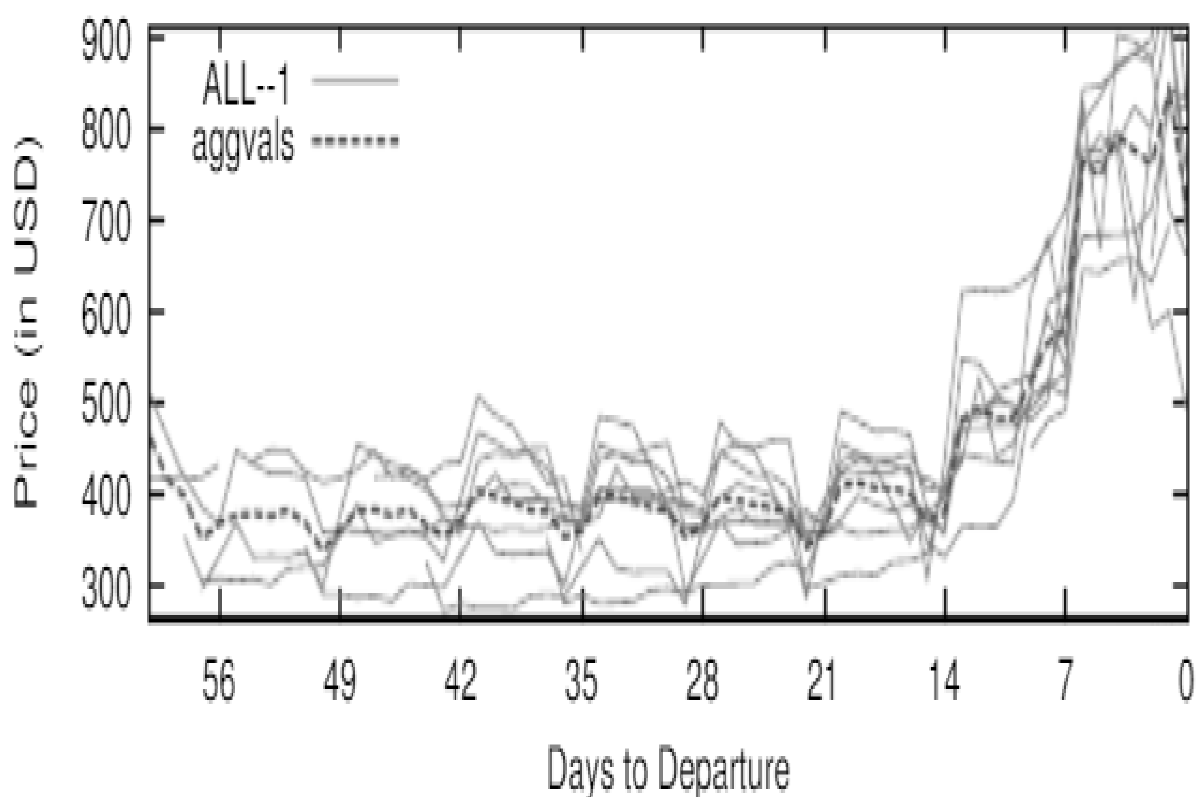


Рис. 2. Графік коливання цін авіаквитків для авіаринку США, розроблений авторами статті (Groves & Gini, 2013)

У статті (Chawla & Kaur, 2017) наведено графік коливання ціни авіаквитків на рейс Делі-Колката за 19 лютого 2017 року. На графіку виділяється різке одноденне (за 15 днів до вильоту) підвищення вартості та плати з 25-ти до 17-ти днів до вильоту (рис. 3).

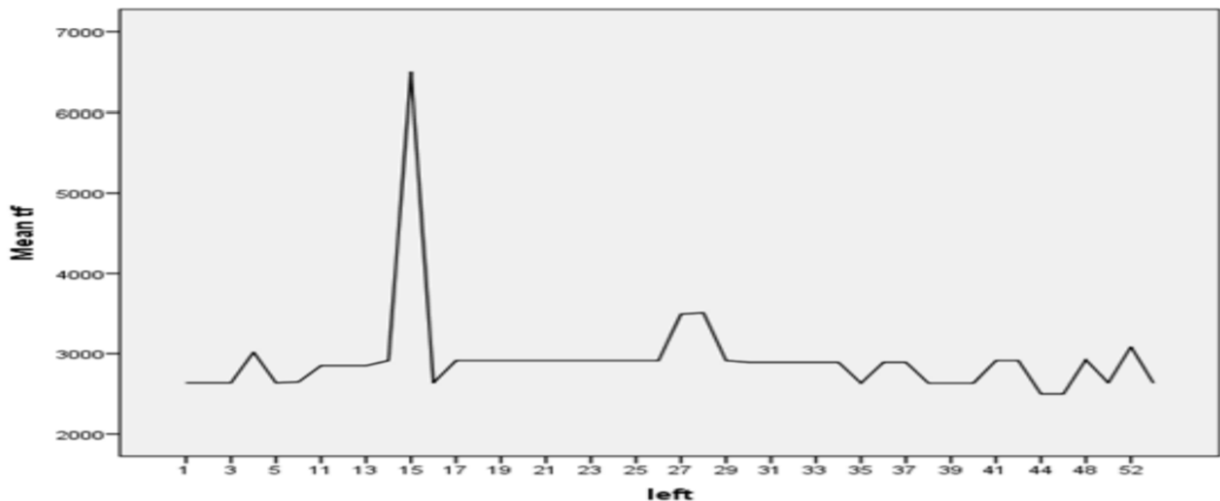


Рис. 3. Графік коливання цін авіаквитків для авіаринку Індії, розроблений авторами статті (Chawla & Kaur, 2017)

Очевидна проблема вибору-покупки білетів: авіакомпанії «граються» цінами на авіаквитки, забезпечуючи баланс між вартістю квитків та наповненістю салону літака; клієнти намагаються вигадати найнижчу ціну; науковці створюють алгоритми передбачення вартості авіаквитків.

Огляд подібних праць

У статті (Tziridis et al., 2017) автор використав вісім моделей машинного навчання: багатошаровий перцептрон (англ. Multilayer Perceptron, MLP), узагальнена регресійна нейронна мережа (англ. Generalized Regression Neural Network), машина екстремального навчання (англ. Extreme Learning Machine, ELM), дерево випадкових лісів регресії (англ. Random Forest Regression Tree), дерево регресії (англ. Regression Tree), беггінг-дерево регресії (англ. Bagging Regression Tree), регресія методу опорних векторів (поліноміальна та лінійна) (англ. Regression Support Vector Machine (Polynomial and Linear)) та лінійна регресія (англ. Linear Regression, LR).

Результатом роботи стали моделі, які передбачали ціну лише на основі двох параметрів: кількості безкоштовного вантажу та кількості днів, що залишились до вильоту. За результатами порівнянь роботи цих моделей найкращою обрано Bagging Regression Tree. Точність передбачення моделі розраховувалась за алгоритмом MSE та досягла 88%.

У статті (Raja et al., 2018) автори проаналізували індійський ринок внутрішніх авіаперевезень та обрали для дослідження шість найпопулярніших рейсів. Для дослідження було використано моделі логістичної регресії (англ. Logistic Regression), дерево регресії, беггінг-дерево регресії, регресію методу опорних векторів (поліноміальна та лінійна) та лінійну регресію. Навчання моделей відбувалось на відкритих даних, які автори збирали власноруч із сайту expedia.com. Мета моделей – надання порад для найкращого часу купівлі авіаквитка.

Найкращі результати показали моделі Bagging Regression Tree – 87.42% точності передбачення та Regression Tree – 84,13%. Однак, дані моделі потребують багато даних для роботи: дата та час вильоту, дата та час прильоту, тип літака та його номер, назва авіакомпанії, кількість пересадок та поточна ціна квитка.

У статті (Wang et al., 2019) автори створили фреймворк, що використовує публічні дані з малим набором полів та не залежить від певного ринку. Найкращий метод передбачення було обрано експериментально серед декількох алгоритмів

машинного навчання: лінійної регресії, методів опорних векторів (англ. Support Vector Machine – SVM) та випадкового лісу (англ. Random Forests – RF), багатошарового персептрону (англ. Multilayer Perceptron, MLP) та дерев градієнтного бустингу (XGBoost Tree). Оцінювання моделей проводилось алгоритмом середньоквадратичного відхилення (англ. Root Mean Square Error, RMSE) та налаштованого коефіцієнта детермінації (R^2_{adj}).

Найкращим методом із результатом R^2_{adj} = 87% було обрано метод випадкового лісу. Цей алгоритм робить передбачення, використовуючи такі значущі параметри набору даних: відстань (рівень значущості 51%), клас місця (27%), кількість пасажирів даного рейсу (6%), коефіцієнт завантаження (4%) та коефіцієнт конкурентоздатності (4%).

У статті (Rajankar et al., 2019) автор порівнював методи SVM, дерев прийняття рішень, метод k-найближчих сусідів (англ. k-nearest neighbors, KNN), Bagging Tree, випадкового лісу та лінійної регресії. Оцінка результатів проводилась методом коефіцієнта детермінації (R^2). Найкращі результати показали методи випадкового лісу та дерев прийняття рішень – 68% та 67% відповідно.

У статті (Chawla & Kaur, 2017) автори обрали дані шести внутрішньо індійських рейсів за 60 днів та порівняли регресійні (SVM, випадковий ліс, градієнтний бустинг, AdaBoost, метод k-найближчих сусідів) та класифікаційні (SVM, випадковий ліс, AdaBoost, наївний баєсів класифікатор, KNN, логістична регресія) моделі. Найвищу точність у 84,01% показав наївний баєсів класифікатор та SVM – 83,97% як класифікаційна модель та 83,27% як регресійна модель.

Взявши за основу власні дослідження, автори розробили прототип комп'ютерного додатка, однак зіткнулись із обмеженням: для коректної роботи програмі необхідні дані авіаційного пального. Як наслідок, застосунок повинен спочатку спрогнозувати вартість пального (автори обрали метод SVM, який показав найвищу точність), а потім – ціну авіаквитка.

Висновки

У цій статті наведено огляд праць, автори яких порівнювали моделі машинного навчання. Результати праць науковців вказують, що не існує деякого універсального алгоритму для усього ринку пасажирських авіаперевезень, який із високою (> 85%) точністю передбачав би вартість авіаквитка. Досягнення даного напрямку роботи обмежуються лише прямими рейсами певного внутрішнього ринку (США, Індія) та 88% точності прогнозу (Tziridis et al., 2017).

Найкращим результатом можна вважати Bagging Regression Tree-модель, описану в статті (Tziridis et al., 2017). Ця навчена модель може робити передбачення лише на основі двох параметрів: кількості безкоштовного вантажу та кількості днів, що залишились до вильоту. Вона має точність у 88%.

Література

1. Сериков, П. (2019). *Как купить авиабилет максимально дешево или ударим мониторингом по динамическому ценообразованию*. Отримано з https://habr.com/ru/post/460213/
2. Abdella, J. A., Zaki, N., Shuaib, K., & Khan, F. (2019). *Airline ticket price and demand prediction: A survey*. Journal of King Saud University - Computer and Information Sciences. DOI: 10.1016/j.jksuci.2019.02.001.
3. Chawla, B., Kaur, C. (2017). *Airfare Analysis And Prediction Using Data Mining And Machine Learning*. International Journal of Engineering Science Invention, 6(11), 7.
4. Groves, W., & Gini, M. (2013). *An agent for optimizing airline ticket purchasing*. Proceedings of the 2013 international conference on Autonomous agents and multi-agent systems, 1341–1342.
5. Raja, V., Vakil, J., Shah, Y., & Relan, S. (2018). *Prediction of Airfare Using Machine Learning*. International Journal of Science & Engineering Development Research, 3(4), 256-259.
6. Rajankar, S., Sakharkar, N., & Rajankar, O. (2019). *Predicting The Price Of A Flight Ticket With The Use Of Machine Learning Algorithms.* International Journal of Scientific & Technology Research, 8(12), 4.

7. Shy, O. (2008). *How to Price: A Guide to Pricing Techniques and Yield Management*. Almatourism, Cambridge University Press. DOI: 10.6092/issn.2036-5195/1894.
8. Tziridis, K., Kalampokas, Th., Papakostas, G. A., & Diamantaras, K. I. (2017). *Airfare prices prediction using machine learning techniques.* 25th European Signal Processing Conference (EUSIPCO), 1036–1039. DOI: 10.23919/EUSIPCO.2017.8081365.
9. Udachny, A. (2016). *Proven the best time to buy airline tickets*. AirHint. Отримано з https://www.airhint.com/articles/best-time-to-buy-airline-tickets.
10. Wang, T., Pouyanfar, S., Tian, H., Tao, Y., Alonso, M., Luis, S., & Chen, S.-C. (2019). *A Framework for Airfare Price Prediction: A Machine Learning Approach*. 2019 IEEE 20th International Conference on Information Reuse and Integration for Data Science, 200-207. DOI: 10.1109/IRI.2019.00041.

References

1. Serikov, P. (2019). Kak kupit aviabilet maksimalno deshevo ili udarim monitoringom po dinamicheskomu tsenoobrazovaniyu. Retrieved from https://habr.com/ru/post/460213/
2. Abdella, J. A., Zaki, N., Shuaib, K., & Khan, F. (2019*). Airline ticket price and demand prediction: A survey.* Journal of King Saud University - Computer and Information Sciences. DOI: 10.1016/j.jksuci.2019.02.001.
3. Chawla, B., Kaur, C. (2017). *Airfare Analysis And Prediction Using Data Mining And Machine Learning.* International Journal of Engineering Science Invention, 6(11), 7.
4. Groves, W., & Gini, M. (2013). *An agent for optimizing airline ticket purchasing*. Proceedings of the 2013 international conference on Autonomous agents and multi-agent systems, 1341–1342.
5. Raja, V., Vakil, J., Shah, Y., & Relan, S. (2018). *Prediction of Airfare Using Machine Learning*. International Journal of Science & Engineering Development Research, 3(4), 256-259.
6. Rajankar, S., Sakharkar, N., & Rajankar, O. (2019). *Predicting The Price Of A Flight Ticket With The Use Of Machine Learning Algorithms.* International Journal of Scientific & Technology Research, 8(12), 4.
7. Shy, O. (2008). *How to Price: A Guide to Pricing Techniques and Yield Management*. Almatourism, Cambridge University Press. DOI: Tziridis, K., Kalampokas, Th., Papakostas, G. A., & Diamantaras, K. I. (2017). *Airfare prices prediction using machine learning techniques.* 25th European Signal Processing Conference (EUSIPCO), 1036–1039. DOI: 10.23919/EUSIPCO.2017.8081365.
8. Tziridis, K., Kalampokas, Th., Papakostas, G. A., & Diamantaras, K. I. (2017). *Airfare prices prediction using machine learning techniques.* 25th European Signal Processing Conference (EUSIPCO), 1036–1039. DOI: 10.23919/EUSIPCO.2017.8081365.
9. Udachny, A. (2016). *Proven the best time to buy airline tickets*. AirHint. Retrieved from https://www.airhint.com/articles/best-time-to-buy-airline-tickets.
10. Wang, T., Pouyanfar, S., Tian, H., Tao, Y., Alonso, M., Luis, S., & Chen, S.-C. (2019). *A Framework for Airfare Price Prediction: A Machine Learning Approach*. 2019 IEEE 20th International Conference on Information Reuse and Integration for Data Science, 200-207. DOI: 10.1109/IRI.2019.00041.

Стаття надійшла до редакції 26.06.2020
Після доробки 10.08.2020

